

WHAT IS CLAIMED IS:

1. A media and gesture recognition method using a computer system, the method comprising:

viewing and generating a digital representation of a printed media using an electronic visual sensor during a first interaction session;

identifying the printed media using the digital representation of the printed media;

retrieving information corresponding to the viewed printed media from a computer system database;

using the electronic visual sensor to view at least a first gesture of a user relative to at least a portion of the printed media;

interpreting the first finger gesture as a first command; and

based at least in part on the first gesture and the retrieved information, providing at least a portion of the retrieved information.

2. The method as defined in Claim 1, wherein identifying the printed media further comprises recognizing visual features that correspond to scale-invariant features (SIFT).

3. The method as defined in Claim 1, wherein the electronic visual sensor is mounted on a robot, wherein the robot positions itself so as to adequately view the printed media.

4. The method as defined in Claim 1, wherein the electronic visual sensor is automatically tilted to improve the viewing of the printed media.

5. The method as defined in Claim 1, further comprising performing gesture calibration.

6. The method as defined in Claim 1, further comprising performing color balancing calibration based at least in part on a viewed portion of a userhand.

7. The method as defined in Claim 1, further comprising instructing the user to perform at least one gesture during a calibration operation.

8. The method as defined in Claim 1, wherein the first gesture is a diagonal sweep of a fingertip across a page of the printed media.

9. The method as defined in Claim 1, wherein the first gesture is a movement of a fingertip beneath at least a first word.

10. The method as defined in Claim 1, wherein the first gesture is a finger tapping movement.

11. The method as defined in Claim 1, wherein the portion of the retrieved information is a word from the printed media.

12. The method as defined in Claim 1, wherein the portion of the retrieved information is a sentence from the printed media.

13. The method as defined in Claim 1, wherein the portion of the retrieved information is a title of the printed media.

14. The method as defined in Claim 1, wherein the portion of the retrieved information is a table contents corresponding to the printed media.

15. The method as defined in Claim 1, wherein the portion of the retrieved information includes a definition retrieved from an electronic dictionary.

16. The method as defined in Claim 1, wherein the printed media is one of a book, a magazine, a musical score, and a map.

17. The method as defined in Claim 1, further comprising:

detecting an exception condition caused by an inadequate view of the printed media; and

providing the user with instructions on handling the printed media to correct the exception condition.

18. The method as defined in Claim 1, further comprising:

determining that the printed media is inadequately viewed; and
instructing the user to rotate the printed media.

19. The method as defined in Claim 1, further comprising:

detecting a timeout condition; and

based at least in part on detecting the timeout condition, informing the user that the first interaction session is ended.

20. The method as defined in Claim 1, wherein the database includes a preference that controls user interaction with the printed media at least at a book-level and a page-level, and a mapping of regions of the printed media with corresponding actions.

21. The method as defined in Claim 1, further comprising detecting the first gesture by comparing at least a first image and a second image electronic received by the visual sensor.

22. The method as defined in Claim 1, wherein the visual sensor includes at least one of CCD imager, a CMOS imager, and an infrared imager.

23. A vision-based method of processing user interaction with printed media, the method comprising:

- receiving at a computer system a digital representation of a first image of a printed media, wherein the first image was obtained from a first imaging device;

- based at least in part on the digital representation of the first image, retrieving corresponding information from a database;

- receiving a first digital representation of a first image of a user gesture relative to at least a portion of the printed media;

- interpreting the first digital representation of an image of a user gesture; and

- based at least in part on the interpretation of the user gesture and the retrieved database information, providing at least a portion of the retrieved information to the user.

24. The method as defined in Claim 23, wherein interpreting the digital representation of an image of a user gesture further comprises:

- finding averages for corresponding blocks within the first digital representation of the first image of the user gesture;

- subtracting the averages from averages of a prior digital representation of an image to generate a difference matrix having difference blocks;

- discarding difference blocks having averages beneath a first predetermined threshold; and

- discarding difference blocks having averages above a second predetermined threshold.

25. The method as defined in Claim 23, wherein the user gesture is used to select printed media text and wherein providing at least a portion of the retrieved information to the user includes reading aloud the selected text.

26. The method as defined in Claim 23, wherein the user gesture is used to select a printed image in the printed media and wherein providing at least a portion of the retrieved information to the user includes displaying a video related to the printed image.

27. The method as defined in Claim 23, wherein the user gesture is used to select a map location in the printed media, and wherein providing at least a portion of the retrieved information to the user includes providing information related to geographical location correspond to the selected map location.

28. The method as defined in Claim 23, wherein the user gesture is used to select a portion of a musical score in the printed media, and wherein providing at least a portion of the retrieved information to the user includes audibly playing the selected portion of the musical score.

29. The method as defined in Claim 23, wherein the first imaging device is mounted on an autonomous mobile apparatus, the method further comprising automatically positioning the autonomous mobile apparatus based on at least one image of the printed media.

30. The method as defined in Claim 23, further comprising performing lighting calibration.

31. The method as defined in Claim 23, further comprising providing the user with one or more audible media interaction prompts.

32. The method as defined in Claim 23, further comprising:
providing the user with a first prompt;
waiting a first amount of time for the user to respond to the first prompt; and
performing a timeout process if the user does not respond within the first amount of time.

33. The method as defined in Claim 23, further comprising:
determining if the printed media is skewed; and
providing the user with skew correction prompts.

34. The method as defined in Claim 23, further comprising:
determining if the printed media is moving; and
providing the user with an instruction to stop moving the media.
35. The method as defined in Claim 23, further comprising:
determining if at least a first page of the printed media is not within a first image frame; and
informing the user that the system cannot view the entire page.
36. A computer-based printed media interaction apparatus, the apparatus comprising:
an image sensor, the image sensor configured to view printed media;
a database including a mapping of regions of the printed media with corresponding actions;
a gesture tracking module that tracks a user gesture position relative to the printed media based at least in part on images from the image sensor; and
an interaction module that, based at least in part on the user gesture position and database information, provides at least a portion of the database information to the user.
37. The apparatus as defined in Claim 36, further comprising a plurality of motorized wheels under computer control used to position the image sensor to view the printed media.
38. The apparatus as defined in Claim 36, further comprising an exception module that informs the user when the printed media is not being adequately viewed by the image sensor.
39. The apparatus as defined in Claim 36, further comprising an exception module that informs the user when the printed media is moved.
40. The apparatus as defined in Claim 36, wherein the gesture tracking module determines a difference between at least two images and filters out difference values greater than a first amount and difference values less than a second amount.
41. The apparatus as defined in Claim 36, wherein the image sensor is a pan and scan camera.

42. The apparatus as defined in Claim 36, wherein the gesture tracking module determines if the user is making at least one of a read a word gesture and a read a page gesture.

43. The apparatus as defined in Claim 36, wherein the gesture tracking module determines if the gesture corresponds to a request for a word definition.

44. The apparatus as defined in Claim 36, further comprising a dictionary.

45. The apparatus as defined in Claim 36, further comprising a topic-specific dictionary.

46. The apparatus as defined in Claim 36, further comprising a network link to information corresponding to the printed media.

47. The apparatus as defined in Claim 36, further comprising a speaker that audibly provides the database information to the user.

48. The apparatus as defined in Claim 36, further comprising a display that visually provides the database information to the user.

49. The apparatus as defined in Claim 36, wherein the printed media is one of a magazine, a musical score, and a book.

50. The apparatus as defined in Claim 36, further comprising a character recognition module that converts images of text into text.

51. A media and gesture recognition apparatus, the apparatus comprising:

an image sensor that views printed media;

a recognition module that identifies the printed media based on image information from the image sensor;

a database that stores information that relates portions of the printed media with corresponding actions;

a gesture tracking module that identifies user gestures relative to the printed media based at least in part on images from the image sensor; and

an interaction module that, based at least in part on the user gesture and database information, provides at least a portion of the database information to the user.

52. The apparatus as defined in Claim 51, wherein the apparatus is stationary.

53. The apparatus as defined in Claim 51, wherein the apparatus includes computer controlled motors that move the apparatus to view the printed media.

54. The apparatus as defined in Claim 51, further comprising a print media support apparatus.

55. The apparatus as defined in Claim 51, wherein the database includes text from the printed media, the apparatus further comprising a speaker that audibly reads at least a portion of the text to the user.

56. The apparatus as defined in Claim 51, further comprising a character recognition module that converts images of text into text.

57. The apparatus as defined in Claim 51, further comprising a dictionary.